The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

An Intuitive Introduction to Targeted Maximum Likelihood Estimation (TMLE)

SSE 708: Machine Learning in the Era of Big Data

David McCoy PhD, MSc

Division of Biostatistics, UC Berkeley
Edward's Lifesciences

July 7-11, 2025

Table of Contents

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

1 The Causal Question

2 Traditional Approaches & Their Limits

3 TMLE: A Smarter, Targeted Approach

4 The Theory Behind the Targeting Step

5 Summary

6 TMLE for Longitudinal Data

7 Impact and Future Directions

Let's Start with a Question

Imagine you are a public health researcher. You want to know:

Motivating Public Health Question

Does engaging in regular physical activity **cause** a reduction in 5-year mortality among the elderly?

- **Treatment (A):** Engaging in high levels of leisure-time physical activity (LTPA).
- **Outcome (Y):** Death within 5 years.
- **Our Goal:** Estimate the **Average Treatment Effect (ATE)**, or Risk Difference.

This is a causal question. We want to know what would happen if we could intervene.

The Gold Standard: The Randomized Trial (RCT)

If we had unlimited resources and no ethical constraints, we would run a perfect experiment:

1. Take a large group of elderly individuals.
2. **Randomly assign** half to a mandatory exercise program ($A = 1$) and half to a no-exercise program ($A = 0$).
3. Follow everyone for 5 years and compare the death rates.

Randomization is key! It ensures that the *only* systematic difference between the groups at the start is the treatment.

The Problem

We can't do this! We have **observational data**, not experimental data.

The Challenge of Observational Data: Confounding

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

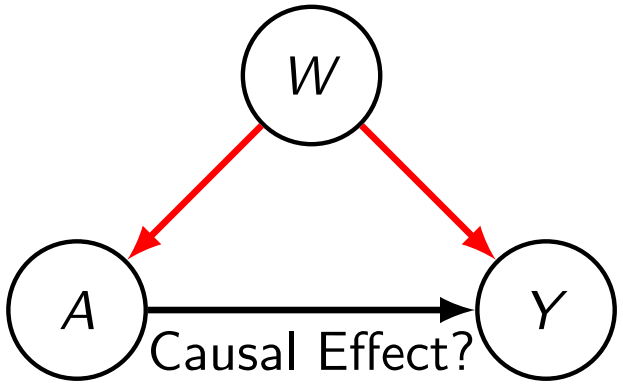
Impact and Future Directions

In the real world, people who exercise might be different from people who don't in many ways.

These other factors, which we call **confounders (W)**, can distort the relationship between our treatment and outcome.

- Age
- Pre-existing illness
- Smoking status
- Self-rated health

This mixing of effects is called **confounding**.



The confounder (W) affects both treatment (A) and outcome (Y), creating a "backdoor path" of association.

Our task: Find a way to estimate the causal effect of $A \rightarrow Y$ as if we had done an RCT, by using our data on A, Y, and W.

Approach 1: Outcome Regression (G-Computation)

The Idea

Use statistical modelling to "adjust" for confounders.

1. **Model the outcome:**
   fit a regression (e.g. logistic) for $P(Y = 1 \mid A, W)$.

$$P(Y = 1 \mid A, W) \approx \text{Model}(\beta_0 + \beta_1 A + \beta_2 W).$$

2. **Play "what if":**
   - predict each person's outcome if $A = 1$ (had exercised);
   - predict again if $A = 0$ (had *not* exercised).
3. **Compare averages:**
   difference of the two mean predictions $\Rightarrow$ estimated causal effect.

The Pitfall

Misspecifying the regression (e.g. omitting an interaction) yields **bias**.

Approach 2: Propensity Scores (IPTW)

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The Idea

Model the *treatment-assignment process* to build a weighted "pseudo–RCT."

1. **Model the treatment:**
   estimate the propensity score $g(A \mid W)$.
2. **Weight the data:**
   each subject gets weight $1/g(A \mid W)$ (large when treatment was unlikely).
3. **Compute a weighted mean:**
   compare treated vs. untreated outcomes in this weighted sample.

The Pitfall

Misspecifying g yields **bias**; extreme scores inflate weights and destabilise estimates.

The Double-Robust Idea: Two Chances to Succeed

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Two Working Models

1. **Outcome model**

$$Q(A, W) = \mathbb{E}[\, Y \mid A, W \,]$$

2. **Propensity-score model**

$$g(W) = P(A = 1 \mid W) \quad \text{and} \quad 1 - g(W) = P(A = 0 \mid W)$$

Key Principle

If *either* model is correct (not necessarily both), we can still obtain a consistent estimate of the Average Treatment Effect.

Augmented IPTW: A Doubly-Robust Estimator

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Estimator for the risk difference $\psi = \mathbb{E}[Y^1 - Y^0]$

$$\widehat{\psi}_{\text{AIPTW}} = \frac{1}{n}\sum_{i=1}^{n}\left[\underbrace{\hat{Q}(1, W_i) - \hat{Q}(0, W_i)}_{\text{plug-in}} + \underbrace{\left\{\frac{A_i}{\hat{g}(W_i)} - \frac{1 - A_i}{1 - \hat{g}(W_i)}\right\}\left(Y_i - \hat{Q}(A_i, W_i)\right)}_{\text{IPW bias-correction}}\right].$$

Double robustness: consistent if *either* $\hat{Q}$ *or* $\hat{g}$ is correct; semiparametrically efficient when *both* are correct.

Why Do We Need More Than A-IPTW?

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Double robustness is great! But A-IPTW has some limitations:

- It's an "estimating equation" estimator, not a plug-in estimator. It doesn't give you an updated probability model, just a final number.
- It can sometimes produce estimates that are nonsensical (e.g., predict a risk of 110%).

The TMLE Insight

What if we could start with an outcome model (Q) and then **cleverly update it** using information from the treatment model (g) to make it **less biased** for the specific question we're asking?

This is exactly what TMLE does. It "targets" the estimation process towards our parameter of interest (the ATE).

The TMLE Recipe: An Intuitive Walk-Through

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

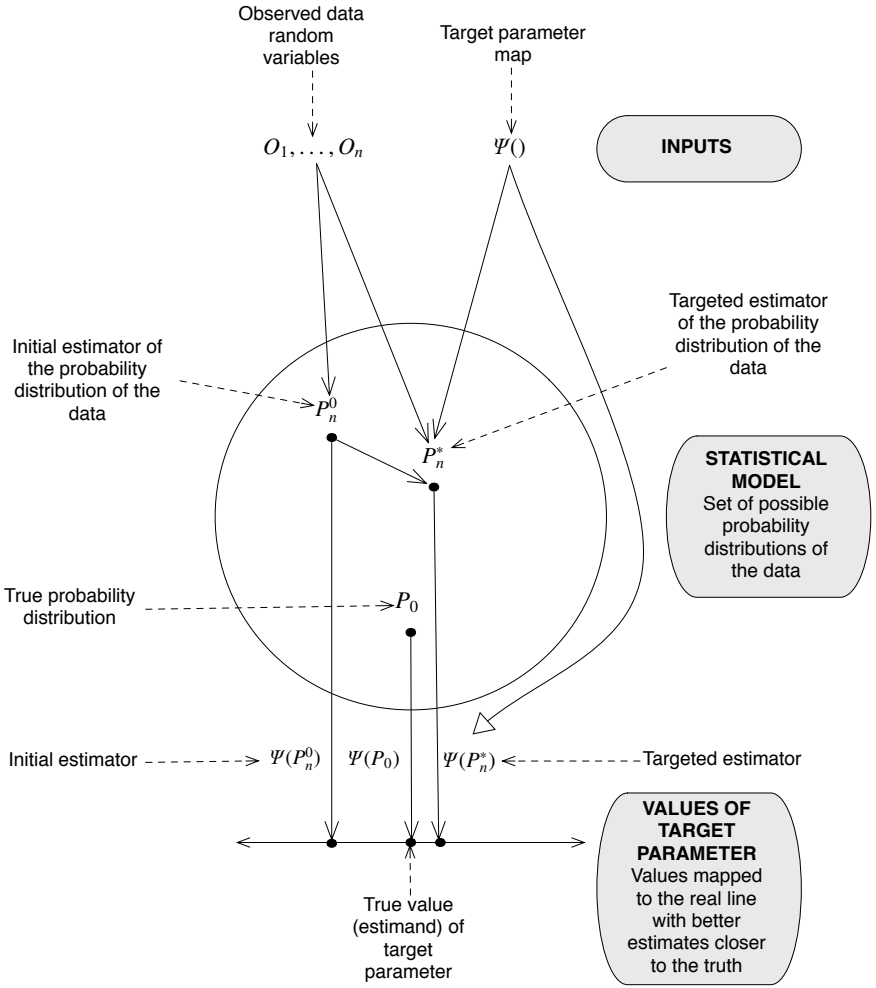
TMLE for Longitudinal Data

Impact and Future Directions

Goal: obtain the best possible plug-in estimate of the ATE.

1. **Initial Q fit**. Use Super Learner to estimate $Q_n^0(A, W) = \widehat{\mathbb{E}}[Y \mid A, W]$.
2. **Fit g**. Estimate the propensity score $g_n(A \mid W)$ via machine learning.
3. **Targeting step**. Update Q_n^0 to Q_n^*:
   - "Fluctuate" Q_n^0 along the *clever covariate* derived from g_n.
   - Find the optimum shift $\hat{\varepsilon}$ by MLE.
4. **Compute the ATE**.

$$\psi_{\text{TMLE}} = \frac{1}{n}\sum_{i=1}^{n}\left(Q_n^*(1, W_i) - Q_n^*(0,\right.$$



The Core Challenge: From Prediction to Causation

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

We begin with a powerful initial estimate for the outcome model, Q_n^0, from Super Learner.

The Problem

Super Learner is built to be an excellent **prediction** machine. It minimizes overall error for $E[Y \mid A, W]$.

The Key Insight

This is **not** the same goal as minimizing the bias for our specific **causal parameter**, the Average Treatment Effect (ψ).

So, how do we intelligently tweak our prediction model to give us a causal answer?

The Core Challenge: From Prediction to Causation

We begin with a powerful initial estimate for the outcome model, Q_n^0, from Super Learner.

The Problem

Super Learner is built to be an excellent **prediction** machine. It minimizes overall error for $E[Y \mid A, W]$.

The Key Insight

This is **not** the same goal as minimizing the bias for our specific **causal parameter**, the Average Treatment Effect (ψ).

So, how do we intelligently tweak our prediction model to give us a causal answer?

The Core Challenge: From Prediction to Causation

We begin with a powerful initial estimate for the outcome model, Q_n^0, from Super Learner.

The Problem

Super Learner is built to be an excellent **prediction** machine. It minimizes overall error for $E[Y \mid A, W]$.

The Key Insight

This is **not** the same goal as minimizing the bias for our specific **causal parameter**, the Average Treatment Effect (ψ).

So, how do we intelligently tweak our prediction model to give us a causal answer?

The Core Challenge: From Prediction to Causation

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

We begin with a powerful initial estimate for the outcome model, Q_n^0, from Super Learner.

The Problem

Super Learner is built to be an excellent **prediction** machine. It minimizes overall error for $E[Y \mid A, W]$.

The Key Insight

This is **not** the same goal as minimizing the bias for our specific **causal parameter**, the Average Treatment Effect (ψ).

So, how do we intelligently tweak our prediction model to give us a causal answer?

The TMLE Strategy: A Targeted Journey

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Our goal is to update our initial estimate Q_n^0 to a new one, Q_n^*, that is tailored to our causal question.

The Strategy in Two Steps:

1. Define a clever "path" that starts at our initial estimate (Q_n^0) and travels in the most efficient direction to reduce bias for our parameter ψ.
2. Use Maximum Likelihood Estimation (MLE) to find the perfect stopping point along that path.

This path is called a "submodel." TMLE chooses the smartest possible path.

The TMLE Strategy: A Targeted Journey

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Our goal is to update our initial estimate Q_n^0 to a new one, Q_n^*, that is tailored to our causal question.

The Strategy in Two Steps:

1. Define a clever "path" that starts at our initial estimate (Q_n^0) and travels in the most efficient direction to reduce bias for our parameter ψ.
2. Use Maximum Likelihood Estimation (MLE) to find the perfect stopping point along that path.

This path is called a "submodel." TMLE chooses the smartest possible path.

The TMLE Strategy: A Targeted Journey

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Our goal is to update our initial estimate Q_n^0 to a new one, Q_n^*, that is tailored to our causal question.

The Strategy in Two Steps:

1. Define a clever "path" that starts at our initial estimate (Q_n^0) and travels in the most efficient direction to reduce bias for our parameter ψ.
2. Use Maximum Likelihood Estimation (MLE) to find the perfect stopping point along that path.

This path is called a "submodel." TMLE chooses the smartest possible path.

The TMLE Strategy: A Targeted Journey

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Our goal is to update our initial estimate Q_n^0 to a new one, Q_n^*, that is tailored to our causal question.

The Strategy in Two Steps:

1. Define a clever "path" that starts at our initial estimate (Q_n^0) and travels in the most efficient direction to reduce bias for our parameter ψ.
2. Use Maximum Likelihood Estimation (MLE) to find the perfect stopping point along that path.

This path is called a "submodel." TMLE chooses the smartest possible path.

What Makes a Path "Smart"?

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Imagine you're at your initial estimate on a map, and the true causal effect is a landmark you want to reach.

- There are infinite paths you could take.
- The "smartest" path is the one that heads most directly toward your landmark.

The Least Favorable Submodel

In statistics, this "smartest path" is called the **Least Favorable Submodel**. It's the direction where our causal parameter, $\psi(Q)$, is **most sensitive** to a small change in our model Q.

By traveling along this path, we get the most "bang for our buck" in reducing bias.

What Makes a Path "Smart"?

Imagine you're at your initial estimate on a map, and the true causal effect is a landmark you want to reach.

- There are infinite paths you could take.
- The "smartest" path is the one that heads most directly toward your landmark.

The Least Favorable Submodel

In statistics, this "smartest path" is called the **Least Favorable Submodel**. It's the direction where our causal parameter, $\psi(Q)$, is **most sensitive** to a small change in our model Q.

By traveling along this path, we get the most "bang for our buck" in reducing bias.

What Makes a Path "Smart"?

Imagine you're at your initial estimate on a map, and the true causal effect is a landmark you want to reach.

- There are infinite paths you could take.
- The "smartest" path is the one that heads most directly toward your landmark.

The Least Favorable Submodel

In statistics, this "smartest path" is called the **Least Favorable Submodel**. It's the direction where our causal parameter, $\psi(Q)$, is **most sensitive** to a small change in our model Q.

By traveling along this path, we get the most "bang for our buck" in reducing bias.

The Clever Covariate: Our GPS

How do we know the direction of this "smartest path"?

The Clever Covariate Defines the Path

The clever covariate, $H(A, W)$, is the mathematical map for the Least Favorable Submodel.

$$H(A, W) = \left(\frac{I(A = 1)}{g_n(1 \mid W)} - \frac{I(A = 0)}{g_n(0 \mid W)} \right)$$

It uses information from the treatment model (g_n, the propensity score) to tell the outcome model (Q_n^0) exactly how to update itself to reduce bias for the causal effect.

The Clever Covariate: Our GPS

How do we know the direction of this "smartest path"?

The Clever Covariate Defines the Path

The clever covariate, $H(A, W)$, is the mathematical map for the Least Favorable Submodel.

$$H(A, W) = \left(\frac{I(A = 1)}{g_n(1 \mid W)} - \frac{I(A = 0)}{g_n(0 \mid W)} \right)$$

It uses information from the treatment model (g_n, the propensity score) to tell the outcome model (Q_n^0) exactly how to update itself to reduce bias for the causal effect.

The Clever Covariate: Our GPS

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

How do we know the direction of this "smartest path"?

The Clever Covariate Defines the Path

The clever covariate, $H(A, W)$, is the mathematical map for the Least Favorable Submodel.

$$H(A, W) = \left(\frac{I(A = 1)}{g_n(1 \mid W)} - \frac{I(A = 0)}{g_n(0 \mid W)} \right)$$

It uses information from the treatment model (g_n, the propensity score) to tell the outcome model (Q_n^0) exactly how to update itself to reduce bias for the causal effect.

Step 1 of Targeting: The Fluctuation Model

We now create a simple, one-parameter logistic regression model to perform the update.

$$\text{logit}\left(Q_n(\epsilon)\right) = \text{logit}\left(Q_n^0\right) + \epsilon \cdot H(A, W)$$

Let's break this down:

- **The Offset**: $\text{logit}(Q_n^0)$ is our fixed starting point.
- **The Direction**: $H(A, W)$ is our clever covariate, telling us which way to go.
- **The Distance**: ϵ is just a single number we need to find, telling us how far to go in that direction.

Step 1 of Targeting: The Fluctuation Model

We now create a simple, one-parameter logistic regression model to perform the update.

$$\text{logit}\left(Q_n(\epsilon)\right) = \text{logit}\left(Q_n^0\right) + \epsilon \cdot H(A, W)$$

Let's break this down:

- **The Offset**: $\text{logit}(Q_n^0)$ is our fixed starting point.
- **The Direction**: $H(A, W)$ is our clever covariate, telling us which way to go.
- **The Distance**: ϵ is just a single number we need to find, telling us how far to go in that direction.

Step 1 of Targeting: The Fluctuation Model

We now create a simple, one-parameter logistic regression model to perform the update.

$$\text{logit}\left(Q_n(\epsilon)\right) = \text{logit}\left(Q_n^0\right) + \epsilon \cdot H(A, W)$$

Let's break this down:

- **The Offset**: $\text{logit}(Q_n^0)$ is our fixed starting point.
- **The Direction**: $H(A, W)$ is our clever covariate, telling us which way to go.
- **The Distance**: ϵ is just a single number we need to find, telling us how far to go in that direction.

Step 1 of Targeting: The Fluctuation Model

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

We now create a simple, one-parameter logistic regression model to perform the update.

$$\text{logit}\left(Q_n(\epsilon)\right) = \text{logit}\left(Q_n^0\right) + \epsilon \cdot H(A, W)$$

Let's break this down:

- **The Offset**: $\text{logit}(Q_n^0)$ is our fixed starting point.
- **The Direction**: $H(A, W)$ is our clever covariate, telling us which way to go.
- **The Distance**: ϵ is just a single number we need to find, telling us how far to go in that direction.

Step 2 of Targeting: The MLE Update

We have a simple model with only one unknown parameter: ϵ.
How do we find its best value?

We use Maximum Likelihood Estimation! We ask the data: "What value of ϵ makes the observed outcomes (Y) most likely, given our model?"

The MLE fit gives us our optimal distance, ϵ_n.
Our final, **targeted outcome model** is then simply:
$Q_n^* = Q_n(\epsilon_n)$.
This specific, two-step procedure is what makes TMLE doubly robust and statistically efficient.

Step 2 of Targeting: The MLE Update

We have a simple model with only one unknown parameter: ϵ.
How do we find its best value?

We use Maximum Likelihood Estimation! We ask the data: "What value of ϵ makes the observed outcomes (Y) most likely, given our model?"

The MLE fit gives us our optimal distance, ϵ_n.
Our final, **targeted outcome model** is then simply:
$Q_n^* = Q_n(\epsilon_n)$.
This specific, two-step procedure is what makes TMLE doubly robust and statistically efficient.

Step 2 of Targeting: The MLE Update

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

We have a simple model with only one unknown parameter: ϵ.
How do we find its best value?

We use Maximum Likelihood Estimation! We ask the data: "What value of ϵ makes the observed outcomes (Y) most likely, given our model?"

The MLE fit gives us our optimal distance, ϵ_n.

Our final, **targeted outcome model** is then simply: $Q_n^* = Q_n(\epsilon_n)$.

This specific, two-step procedure is what makes TMLE doubly robust and statistically efficient.

Step 2 of Targeting: The MLE Update

We have a simple model with only one unknown parameter: ϵ.
How do we find its best value?

We use Maximum Likelihood Estimation! We ask the data: "What value of ϵ makes the observed outcomes (Y) most likely, given our model?"

The MLE fit gives us our optimal distance, ϵ_n.
Our final, **targeted outcome model** is then simply:
$Q_n^* = Q_n(\epsilon_n)$.
This specific, two-step procedure is what makes TMLE doubly robust and statistically efficient.

What is the MLE for ϵ *Actually* Doing?

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

We have our simple model: $\text{logit}(Q_n(\epsilon)) = \text{logit}(Q_n^0) + \epsilon \cdot H(A, W)$
When we ask our software to find the Maximum Likelihood Estimate (MLE) for ϵ, we are asking it to do one thing:

The Goal of the Fit

Find the single value of ϵ that makes our observed outcomes (Y) most probable, given our starting point (Q_n^0) and the direction provided by the clever covariate ($H(A, W)$).

In essence, the MLE fit for ϵ quantifies the exact amount of "correction" our initial model needs, but *only in the specific direction* of the clever covariate.

What is the MLE for ϵ *Actually* Doing?

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

We have our simple model: $\text{logit}(Q_n(\epsilon)) = \text{logit}(Q_n^0) + \epsilon \cdot H(A, W)$
When we ask our software to find the Maximum Likelihood Estimate (MLE) for ϵ, we are asking it to do one thing:

The Goal of the Fit

Find the single value of ϵ that makes our observed outcomes (Y) most probable, given our starting point (Q_n^0) and the direction provided by the clever covariate ($H(A, W)$).

In essence, the MLE fit for ϵ quantifies the exact amount of "correction" our initial model needs, but *only in the specific direction* of the clever covariate.

What is the MLE for ϵ *Actually* Doing?

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

We have our simple model: $\text{logit}(Q_n(\epsilon)) = \text{logit}(Q_n^0) + \epsilon \cdot H(A, W)$
When we ask our software to find the Maximum Likelihood Estimate (MLE) for ϵ, we are asking it to do one thing:

The Goal of the Fit

Find the single value of ϵ that makes our observed outcomes (Y) most probable, given our starting point (Q_n^0) and the direction provided by the clever covariate ($H(A, W)$).

In essence, the MLE fit for ϵ quantifies the exact amount of "correction" our initial model needs, but *only in the specific direction* of the clever covariate.

Connecting the Update to the Causal Goal

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Here is the most critical insight:

The "Clever Covariate" is Special

The direction $H(A, W)$ was mathematically chosen because it is the path in "model space" along which our causal parameter, ψ, changes the fastest.

Analogy :

- Your initial estimate Q_n^0 is your current location on a mountain.
- The true causal effect ψ_0 is a landmark in a deep valley.
- The clever covariate $H(A, W)$ points you down the **steepest possible path** that leads directly into that specific valley.

Therefore, when MLE finds the best ϵ to improve the model's fit, it is simultaneously pushing our *causal estimate* as far as the data will support along the most efficient path toward the truth.

Connecting the Update to the Causal Goal

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Here is the most critical insight:

The "Clever Covariate" is Special

The direction $H(A, W)$ was mathematically chosen because it is the path in "model space" along which our causal parameter, ψ, changes the fastest.

Analogy :

- Your initial estimate Q_n^0 is your current location on a mountain.
- The true causal effect ψ_0 is a landmark in a deep valley.
- The clever covariate $H(A, W)$ points you down the **steepest possible path** that leads directly into that specific valley.

Therefore, when MLE finds the best ϵ to improve the model's fit, it is simultaneously pushing our *causal estimate* as far as the data will support along the most efficient path toward the truth.

Connecting the Update to the Causal Goal

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Here is the most critical insight:

The "Clever Covariate" is Special

The direction $H(A, W)$ was mathematically chosen because it is the path in "model space" along which our causal parameter, ψ, changes the fastest.

Analogy :

- Your initial estimate Q_n^0 is your current location on a mountain.
- The true causal effect ψ_0 is a landmark in a deep valley.
- The clever covariate $H(A, W)$ points you down the **steepest possible path** that leads directly into that specific valley.

Therefore, when MLE finds the best ϵ to improve the model's fit, it is simultaneously pushing our *causal estimate* as far as the data will support along the most efficient path toward the truth.

The Final Piece: The Efficient Influence Curve (EIC)

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

In statistical theory, the "best" estimator for a parameter is one that solves its **Efficient Influence Curve (EIC)** equation.

The EIC Equation

The goal is to find a model or distribution estimate, P^*, that makes the average of the EIC over our data equal to zero:

$$\frac{1}{n}\sum_{i=1}^{n} D^*(P^*)(O_i) = 0$$

where $D^*(P)$ is the formula for the EIC.

Solving this equation gives you a **doubly-robust** and **statistically efficient** estimate. The question is, how do we solve it?

The Final Piece: The Efficient Influence Curve (EIC)

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

In statistical theory, the "best" estimator for a parameter is one that solves its **Efficient Influence Curve (EIC)** equation.

The EIC Equation

The goal is to find a model or distribution estimate, P^*, that makes the average of the EIC over our data equal to zero:

$$\frac{1}{n}\sum_{i=1}^{n} D^*(P^*)(O_i) = 0$$

where $D^*(P)$ is the formula for the EIC.

Solving this equation gives you a **doubly-robust** and **statistically efficient** estimate. The question is, how do we solve it?

The Final Piece: The Efficient Influence Curve (EIC)

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

In statistical theory, the "best" estimator for a parameter is one that solves its **Efficient Influence Curve (EIC)** equation.

The EIC Equation

The goal is to find a model or distribution estimate, P^*, that makes the average of the EIC over our data equal to zero:

$$\frac{1}{n}\sum_{i=1}^{n} D^*(P^*)(O_i) = 0$$

where $D^*(P)$ is the formula for the EIC.

Solving this equation gives you a **doubly-robust** and **statistically efficient** estimate. The question is, how do we solve it?

The Magic: MLE Solves the EIC Equation!

This is the idea of the TMLE algorithm.
For our specific logistic fluctuation model:
$\text{logit}(Q) = \text{logit}(Q_n^0) + \epsilon \cdot H(A, W)$

The Key Connection: The "score equation" that the logistic regression software solves to find the MLE for ϵ is **mathematically identical** to the Efficient Influence Curve equation for our causal parameter!

$$\sum_{i=1}^{n} \underbrace{(Y_i - Q_n(\epsilon)(A_i, W_i))\, H(A_i, W_i)}_{\text{Score for observation } i} \equiv \sum_{i=1}^{n} \underbrace{D^*(P_n^*)(O_i)}_{\text{EIC for observation } i} = 0$$

The Punchline

When you ask for the MLE of ϵ, you are, without any extra work, commanding the software to solve the EIC equation. This is why the resulting updated model, Q_n^*, yields a doubly-robust and efficient estimate of the causal effect. The fluctuation model was perfectly designed to make this happen.

The Magic: MLE Solves the EIC Equation!

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

This is the idea of the TMLE algorithm.
For our specific logistic fluctuation model:
$\text{logit}(Q) = \text{logit}(Q_n^0) + \epsilon \cdot H(A, W)$

The Key Connection: The "score equation" that the logistic regression software solves to find the MLE for ϵ is **mathematically identical** to the Efficient Influence Curve equation for our causal parameter!

$$\sum_{i=1}^{n} \underbrace{(Y_i - Q_n(\epsilon)(A_i, W_i))\, H(A_i, W_i)}_{\text{Score for observation } i} \equiv \sum_{i=1}^{n} \underbrace{D^*(P_n^*)(O_i)}_{\text{EIC for observation } i} = 0$$

The Punchline

When you ask for the MLE of ϵ, you are, without any extra work, commanding the software to solve the EIC equation. This is why the resulting updated model, Q_n^*, yields a doubly-robust and efficient estimate of the causal effect. The fluctuation model was perfectly designed to make this happen.

The Magic: MLE Solves the EIC Equation!

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

This is the idea of the TMLE algorithm.
For our specific logistic fluctuation model:
$\text{logit}(Q) = \text{logit}(Q_n^0) + \epsilon \cdot H(A, W)$

The Key Connection: The "score equation" that the logistic regression software solves to find the MLE for ϵ is **mathematically identical** to the Efficient Influence Curve equation for our causal parameter!

$$\sum_{i=1}^{n} \underbrace{(Y_i - Q_n(\epsilon)(A_i, W_i))\, H(A_i, W_i)}_{\text{Score for observation } i} \equiv \sum_{i=1}^{n} \underbrace{D^*(P_n^*)(O_i)}_{\text{EIC for observation } i} = 0$$

The Punchline

When you ask for the MLE of ϵ, you are, without any extra work, commanding the software to solve the EIC equation. This is why the resulting updated model, Q_n^*, yields a doubly-robust and efficient estimate of the causal effect. The fluctuation model was perfectly designed to make this happen.

A Deeper Issue: Are Our Models "Well-Behaved"?

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

For our final p-values and confidence intervals to be trustworthy, statistical theory requires our initial estimate, Q_n^0, to be "sufficiently well-behaved."

The Intuition

Think of it like this: if our machine learning model from Super Learner is excessively complex or "spiky," it might fit the data in erratic ways.

This erratic behavior can violate the assumptions needed for our standard statistical theory to apply, potentially leading to incorrect confidence intervals.

A Deeper Issue: Are Our Models "Well-Behaved"?

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

For our final p-values and confidence intervals to be trustworthy, statistical theory requires our initial estimate, Q_n^0, to be "sufficiently well-behaved."

The Intuition

Think of it like this: if our machine learning model from Super Learner is excessively complex or "spiky," it might fit the data in erratic ways.

This erratic behavior can violate the assumptions needed for our standard statistical theory to apply, potentially leading to incorrect confidence intervals.

A Deeper Issue: Are Our Models "Well-Behaved"?

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

For our final p-values and confidence intervals to be trustworthy, statistical theory requires our initial estimate, Q_n^0, to be "sufficiently well-behaved."

The Intuition

Think of it like this: if our machine learning model from Super Learner is excessively complex or "spiky," it might fit the data in erratic ways.

This erratic behavior can violate the assumptions needed for our standard statistical theory to apply, potentially leading to incorrect confidence intervals.

The "Donsker Class" Condition

The technical name for this "well-behaved" property is that the functions produced by our estimator should belong to a **Donsker class**.

Think of it as a "Complexity Speed Limit"

If an algorithm is too flexible (e.g., a very deep neural network), it might be so complex that it "breaks the speed limit," and our statistical inference can no longer be trusted.

This presents a dilemma: We want to use the most powerful machine learning tools to reduce bias, but doing so might invalidate our p-values!

The "Donsker Class" Condition

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The technical name for this "well-behaved" property is that the functions produced by our estimator should belong to a **Donsker class**.

Think of it as a "Complexity Speed Limit"

If an algorithm is too flexible (e.g., a very deep neural network), it might be so complex that it "breaks the speed limit," and our statistical inference can no longer be trusted.

This presents a dilemma: We want to use the most powerful machine learning tools to reduce bias, but doing so might invalidate our p-values!

The "Donsker Class" Condition

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The technical name for this "well-behaved" property is that the functions produced by our estimator should belong to a **Donsker class**.

Think of it as a "Complexity Speed Limit"

If an algorithm is too flexible (e.g., a very deep neural network), it might be so complex that it "breaks the speed limit," and our statistical inference can no longer be trusted.

This presents a dilemma: We want to use the most powerful machine learning tools to reduce bias, but doing so might invalidate our p-values!

The Solution: Cross-Validated TMLE (CV-TMLE)

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Why CV-TMLE?

Out-of-sample predictions break the training–evaluation dependence, so we no longer need each learner to satisfy Donsker conditions.

1. **Partition** the data into K folds: training folds $\mathcal{T}_k$ and validation fold $\mathcal{V}_k$.
2. **For each** $k = 1, \ldots, K$:
   - Fit Super Learner on $\mathcal{T}_k$.
   - Obtain initial $Q^0_{n,-k}(A, W)$ and $g_{n,-k}(A \mid W)$ for subjects in $\mathcal{V}_k$.
   - Run the targeting step on that same fit $\Rightarrow Q^*_{n,-k}$.
3. **Stack** the K validation-fold outputs so every subject now has Q^*, g, and H values that were generated without seeing their own outcome.
4. **Aggregate across folds:**

$$\hat{\psi}_{\text{CV-TMLE}} = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i \in \mathcal{V}_k} \left(Q^*_{n,-k}(1, W_i) - Q^*_{n,-k}(0, W_i) \right),$$

which is algebraically identical to $\frac{1}{n} \sum_{i=1}^{n} (Q^*(1, W_i) - Q^*(0, W_i))$. Influence-curve SEs and CIs use these same cross-validated predictions.

Key takeaway: Every prediction used in targeting was produced by a model that never saw that observation—so we can unleash highly flexible learners without breaking asymptotic theory.

Why CV-TMLE Matters

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

What CV-TMLE Really Solves

- Ordinary TMLE needs the learning algorithms to be *mathematically tame* (so-called "Donsker").
- Cross-validation breaks the link between *fitting* and *evaluating*, giving us **honest, out-of-sample** predictions.
- That single trick removes the "tameness" requirement.

Take-home message for practice:
With CV-TMLE you can load your Super Learner library with any modern ML tool—deep nets, XGBoost, huge forests—and still get valid confidence intervals and p-values.

You still need:

- Reasonable overlap (positivity) in your data, and
- At least one model (Q or g) to learn at a decent rate (roughly faster than $n^{-1/4}$).

But you *don't* have to worry about whether each learner satisfies complicated empirical-process conditions.

Why is TMLE So Powerful?

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

TMLE: The Best of Both Worlds

- **Double Robust:** Just like A-IPTW, it gives a consistent estimate if either the initial outcome model (Q_n^0) or the treatment model (g_n) is consistent.
- **Efficient:** If *both* models are consistent, TMLE is statistically efficient. This means it achieves the lowest possible variance for the given assumptions—your confidence intervals will be as tight as possible.
- **Substitution Estimator:** It produces an actual, updated probability distribution (P_n^*) where predictions are guaranteed to be in the valid range (e.g., probabilities between 0 and 1).
- **Uses Machine Learning:** It naturally incorporates flexible tools like Super Learner for both the Q and g models, minimizing bias from model misspecification.

SPPARCS Example: The TMLE Steps

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

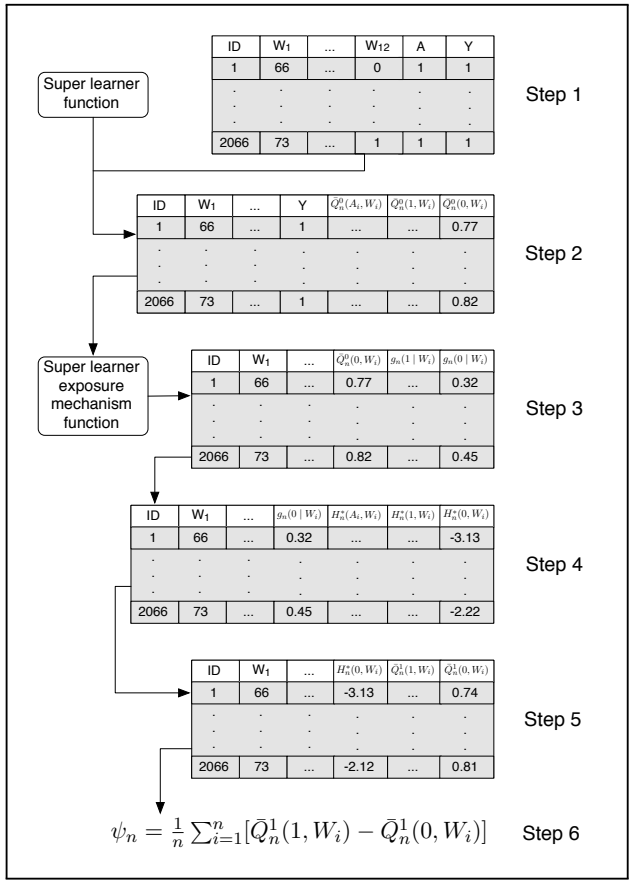
TMLE for Longitudinal Data

Impact and Future Directions

1. **Initial Q_n^0 estimate**. Super Learner (logistic reg., GAMs, etc.) predicts 5-yr mortality from LTPA + 12 confounders.
2. **Estimate g_n**. Super Learner predicts probability of high LTPA given the same 12 confounders.
3. **Targeting step with clever covariate**.

$$H(A, W) = \frac{\mathbf{1}\{A = 1\}}{g_n(1 \mid W)} - \frac{\mathbf{1}\{A = 0\}}{g_n(0 \mid W)}$$

   Regress Y on $H(A, W)$ with logit(Q_n^0) as offset $\rightarrow$ obtain $\hat{\varepsilon}$.
4. **Update to Q_n^***. Apply $\hat{\varepsilon}$ to get targeted mortality predictions.
5. **Calculate final ATE**. Average $Q_n^*(1, W_i) - Q_n^*(0, W_i)$ across all subjects.



SPPARCS Example: Results Interpretation

The TMLE Estimate

The final TMLE estimate for the risk difference was -**0.055**.

Inference:

- The standard error was calculated using the *influence curve*, a robust statistical tool that TMLE provides.
- 95% Confidence Interval: [-0.078, -0.033]
- p-value: < 0.001

Interpretation: After adjusting for a rich set of confounders using a doubly-robust, semi-parametric efficient method, we estimate that meeting or exceeding recommended levels of physical activity **reduces the absolute risk of 5-year mortality by 5.5 percentage points** in this elderly population. This result is statistically significant.

Key Takeaways

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

- Estimating causal effects from observational data is hard because of **confounding**.
- Traditional methods like outcome regression or IPTW rely on a **single model being perfectly correct**, which is a risky assumption.
- **Double-robust** methods give you two chances to get the right answer.
- **TMLE** is a modern, double-robust estimator that:
  - Uses machine learning to avoid restrictive modeling assumptions.
  - Updates an initial estimate in a "targeted" way to reduce bias for your specific question.
  - Is statistically efficient, giving you the most precise estimates and narrowest confidence intervals possible.

Back to the HEARTFIX Simulation

- Earlier you estimated the causal effect of the exercise intervention using **G-Computation** and **IPTW**.
- Now we will re-use the **same simulated dataset** to demonstrate a **manual TMLE update**.
- Goal: show that a few lines of R code can transform your initial Super Learner predictions into a *doubly-robust*, *efficient* estimate of the ATE.

Manual TMLE Update — Code Walk-Through

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

1. **Build the "clever covariate"**

$$H_i = \frac{\mathbb{I}(A_i = 1)}{\hat{g}(1 \mid W_i)} - \frac{\mathbb{I}(A_i = 0)}{1 - \hat{g}(1 \mid W_i)}$$

```
H <- A/gn1 - (1-A)/gn0
```

2. **Define the least-favourable submodel**

$$\text{logit}\big(Q_i(\varepsilon)\big) = \text{logit}\big(\hat{Q}_i^0\big) + \varepsilon\, H_i$$

```
logit <- function(p) log(p/(1-p))
```

3. **Estimate the optimal shift $\hat{\varepsilon}$**

$$\hat{\varepsilon} = \arg\max_{\varepsilon} \sum_{i=1}^{n} \Big[Y_i \log Q_i(\varepsilon) + (1 - Y_i) \log\{1 - Q_i(\varepsilon)\} \Big]$$

```
eps <- coef(glm(Y   -1 + H + offset(logit(QAn)), ...))
```

4. **Update the predictions**

$$\hat{Q}_i^*(1) = \text{logit}^{-1}\big(\text{logit}(\hat{Q}_i^0(1)) + \hat{\varepsilon}/\hat{g}_i(1)\big)$$

$$\hat{Q}_i^*(0) = \text{logit}^{-1}\big(\text{logit}(\hat{Q}_i^0(0)) - \hat{\varepsilon}/\hat{g}_i(0)\big)$$

```
Q1s <- plogis(logit(Q1n) + eps/gn1)
Q0s <- plogis(logit(Q0n) - eps/gn0)
```

5. **Plug-in Average Treatment Effect**

$$\hat{\psi}_{\text{TMLE}} = \frac{1}{n} \sum_{i=1}^{n} \big[\hat{Q}_i^*(1) - \hat{Q}_i^*(0)\big]$$

```
psi <- mean(Q1s - Q0s)
```

Result: a bias-reduced, efficient ATE estimate *and* an updated probability model that respects the [0,1] bounds.

The Evolving World of TMLE

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

TMLE is an active area of research, with many powerful extensions:

- **Collaborative TMLE (C-TMLE):** Uses information from the outcome model to help build a better propensity score model.
- **Cross-Validated TMLE (CV-TMLE):** Uses cross-validation to prevent overfitting during the targeting step, improving performance in smaller samples.
- **TMLE for Complex Data:** Versions exist for longitudinal data, survival analysis, mediation analysis, and more.

The Big Picture

Targeted Learning provides a principled roadmap for combining sophisticated machine learning with rigorous statistical theory to answer the scientific questions we care most about.

The Targeted Learning Ecosystem: Key Applications

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The core TMLE idea has been extended to answer a wide range of scientific questions. Here are a few major areas:

Mediation Analysis

Question: How much of a treatment's effect is *direct*, and how much operates *indirectly* through a mediator? **TMLE's Role:** Provides doubly-robust estimates of natural direct and indirect effects, even with complex interactions.

Optimal Dynamic Treatment Regimes

Question: What is the best sequence of treatments for a patient, where each decision is tailored to their evolving health status? (e.g., "If patient's CD4 count is below X, switch to drug B; otherwise, continue drug A.") **TMLE's Role:** Estimates the mean outcome if everyone in a population followed a specific, complex treatment rule.

Data-Adaptive Target Parameters

Question: Instead of pre-specifying a parameter, can we let the data tell us what scientific question is most supported by the evidence? (e.g., finding the subgroup with the largest treatment effect). **TMLE's Role:** Provides a framework for valid statistical inference even when the research question is adapted based on the data.

The Targeted Learning Ecosystem: Key Applications

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The core TMLE idea has been extended to answer a wide range of scientific questions. Here are a few major areas:

Mediation Analysis

Question: How much of a treatment's effect is *direct*, and how much operates *indirectly* through a mediator? **TMLE's Role:** Provides doubly-robust estimates of natural direct and indirect effects, even with complex interactions.

Optimal Dynamic Treatment Regimes

Question: What is the best sequence of treatments for a patient, where each decision is tailored to their evolving health status? (e.g., "If patient's CD4 count is below X, switch to drug B; otherwise, continue drug A.") **TMLE's Role:** Estimates the mean outcome if everyone in a population followed a specific, complex treatment rule.

Data-Adaptive Target Parameters

Question: Instead of pre-specifying a parameter, can we let the data tell us what scientific question is most supported by the evidence? (e.g., finding the subgroup with the largest treatment effect). **TMLE's Role:** Provides a framework for valid statistical inference even when the research question is adapted based on the data.

The Targeted Learning Ecosystem: Key Applications

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The core TMLE idea has been extended to answer a wide range of scientific questions. Here are a few major areas:

Mediation Analysis

Question: How much of a treatment's effect is *direct*, and how much operates *indirectly* through a mediator? **TMLE's Role:** Provides doubly-robust estimates of natural direct and indirect effects, even with complex interactions.

Optimal Dynamic Treatment Regimes

Question: What is the best sequence of treatments for a patient, where each decision is tailored to their evolving health status? (e.g., "If patient's CD4 count is below X, switch to drug B; otherwise, continue drug A.") **TMLE's Role:** Estimates the mean outcome if everyone in a population followed a specific, complex treatment rule.

Data-Adaptive Target Parameters

Question: Instead of pre-specifying a parameter, can we let the data tell us what scientific question is most supported by the evidence? (e.g., finding the subgroup with the largest treatment effect). **TMLE's Role:** Provides a framework for valid statistical inference even when the research question is adapted based on the data.

The Targeted Learning Ecosystem: Key R Packages

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

This entire ecosystem is supported by a suite of powerful, open-source R packages.

Core Engines

- `sl3`: A modern, object-oriented engine for building Super Learners. Fast, extensible, and the new standard.
- `tmle3`: The modern, unified framework for Targeted Learning. It is designed to be modular, allowing you to define custom parameters, learners, and TMLE procedures with ease.

Legacy Packages

- `SuperLearner`: The original, widely-used Super Learner package.
- `tmle`: The original package for point-treatment TMLE.
- `ltmle`: The original package for Longitudinal TMLE. Still very powerful and widely cited.

Current Best Practice: New projects should use the modern `sl3` and `tmle3` ecosystem.

Specific Applications

- `medshift`: A `tmle3`-based package for complex stochastic mediation analysis.
- `lmtp`: An easy-to-use package for estimating the effects of Longitudinal Modified Treatment Policies.

The Targeted Learning Ecosystem: Key R Packages

This entire ecosystem is supported by a suite of powerful, open-source R packages.

Core Engines

- `sl3`: A modern, object-oriented engine for building Super Learners. Fast, extensible, and the new standard.
- `tmle3`: The modern, unified framework for Targeted Learning. It is designed to be modular, allowing you to define custom parameters, learners, and TMLE procedures with ease.

Legacy Packages

- `SuperLearner`: The original, widely-used Super Learner package.
- `tmle`: The original package for point-treatment TMLE.
- `ltmle`: The original package for Longitudinal TMLE. Still very powerful and widely cited.

Current Best Practice: New projects should use the modern `sl3` and `tmle3` ecosystem.

Specific Applications

- `medshift`: A `tmle3`-based package for complex stochastic mediation analysis.
- `lmtp`: An easy-to-use package for estimating the effects of Longitudinal Modified Treatment Policies.

The Targeted Learning Ecosystem: Key R Packages

This entire ecosystem is supported by a suite of powerful, open-source R packages.

Core Engines

- `sl3`: A modern, object-oriented engine for building Super Learners. Fast, extensible, and the new standard.
- `tmle3`: The modern, unified framework for Targeted Learning. It is designed to be modular, allowing you to define custom parameters, learners, and TMLE procedures with ease.

Legacy Packages

- `SuperLearner`: The original, widely-used Super Learner package.
- `tmle`: The original package for point-treatment TMLE.
- `ltmle`: The original package for Longitudinal TMLE. Still very powerful and widely cited.

Current Best Practice: New projects should use the modern `sl3` and `tmle3` ecosystem.

Specific Applications

- `medshift`: A `tmle3`-based package for complex stochastic mediation analysis.
- `lmtp`: An easy-to-use package for estimating the effects of Longitudinal Modified Treatment Policies.

The Next Level: Time-Varying Treatments

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Many public-health questions involve treatments or exposures that change over time.

Motivating Question

Does **sustained adherence** to a treatment regimen (e.g. HIV therapy) improve a long-term outcome, given that a patient's health both influences *and* is influenced by adherence at each visit?

This introduces a major new challenge: **Time-Varying Confounding**.

The Next Level: Time-Varying Treatments

Many public-health questions involve treatments or exposures that change over time.

Motivating Question

Does **sustained adherence** to a treatment regimen (e.g. HIV therapy) improve a long-term outcome, given that a patient's health both influences *and* is influenced by adherence at each visit?

This introduces a major new challenge: **Time-Varying Confounding**.

The Next Level: Time-Varying Treatments

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Many public-health questions involve treatments or exposures that change over time.

Motivating Question

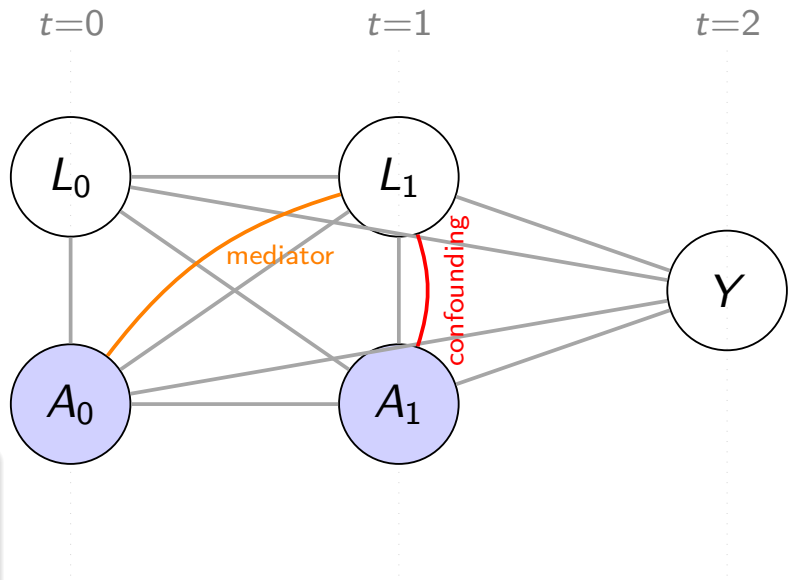
Does **sustained adherence** to a treatment regimen (e.g. HIV therapy) improve a long-term outcome, given that a patient's health both influences *and* is influenced by adherence at each visit?

This introduces a major new challenge: **Time-Varying Confounding**.

The Challenge: Time-Varying Confounding

A two-time-point example

- L_0 Baseline health
- A_0 Treatment at $t = 0$
- L_1 Health at $t = 1$
- A_1 Treatment at $t = 1$
- Y Final outcome ($t = 2$)



Why it's tricky

L_1 is **both**

- a confounder of $A_1 \to Y$, and
- a mediator of $A_0 \to Y$.

Standard regression or IPTW breaks one path while biasing the other.

Why Standard Methods Fail

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Regression (G-Computation)

Adjusting for L_1 blocks the indirect effect of A_0 through L_1, yielding a biased total-effect estimate.

Standard IPTW

Weights based on $P(A_1 \mid L_1, A_0, L_0)$ depend on L_1, which is itself affected by $A_0 \rightarrow$ bias.

We need methods tailored to confounders that lie on the causal path.

Why Standard Methods Fail

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Regression (G-Computation)

Adjusting for L_1 blocks the indirect effect of A_0 through L_1, yielding a biased total-effect estimate.

Standard IPTW

Weights based on $P(A_1 \mid L_1, A_0, L_0)$ depend on L_1, which is itself affected by $A_0 \rightarrow$ bias.

We need methods tailored to confounders that lie on the causal path.

Why Standard Methods Fail

Regression (G-Computation)

Adjusting for L_1 blocks the indirect effect of A_0 through L_1, yielding a biased total-effect estimate.

Standard IPTW

Weights based on $P(A_1 \mid L_1, A_0, L_0)$ depend on L_1, which is itself affected by $A_0 \rightarrow$ bias.

We need methods tailored to confounders that lie on the causal path.

Seq. G-Computation (A): Forward Simulation

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

"Parametric G-formula"

1. Fit models for each time-updated variable $L_t \,|\, \bar{L}_{t-1}, \bar{A}_{t-1}$ and for $Y \,|\, \bar{L}_T, \bar{A}_T$.
2. Draw L_0 from its empirical distribution.
3. For $t = 0, \ldots, T-1$ $\quad \triangleright$ set A_t to the policy value $\quad \triangleright$ simulate L_{t+1}.
4. Repeat many Monte-Carlo paths; average Y.

Pitfall

Misspecify *any* model $\rightarrow$ biased final estimate.

Seq. G-Computation (B): Backward Recursion (ICE)

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

$$\psi = \mathbb{E}\big[Q_0(\bar{a})\big], \qquad \begin{cases} Q_T = \mathbb{E}[Y \mid \bar{A}_T = \bar{a}_T, \bar{L}_T] \\ Q_t = \mathbb{E}[Q_{t+1} \mid \bar{A}_t = \bar{a}_t, \bar{L}_t], \quad t = T-1, \ldots, 0 \end{cases}$$

- Fit the T conditional-expectation models— this is "iterated conditional expectation".
- Plug in policy values $\bar{a}$, then average $Q_0(\bar{a})$ over the baseline covariates.

Same pitfall

Every conditional-mean model must be correct.

Approach 2: Sequential IPTW

Core idea

Weight each subject by the inverse probability of their entire treatment *history*; then fit a simple model in the weighted sample.

$$SW_i = \prod_{t=0}^{T} \frac{\Pr(A_t = a_t)}{\Pr(A_t = a_t \mid \bar{L}_t, \bar{A}_{t-1})}$$

- "Stabilized" numerators keep weights finite.
- Under correct g_t models, the weighted data emulate a sequence of randomised trials.

Pitfall

Needs a correct treatment model at every step; the product can explode if any propensity is near 0 or 1.

The Modern Solution: Longitudinal TMLE (LTMLE)

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

- Combines the G-formula's substitution property with IPTW's robustness.
- Performs a targeted update (TMLE) *at each time point*, working backwards from T to 0.
- Can use Super Learner (plus CV-TMLE) for every Q_t and g_t.

Key insight

The influence-function update at time t removes bias that earlier steps left behind, giving doubly-robust accuracy *iteratively*.

LTMLE Algorithm – High-Level Steps

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

1. **Time** T Fit TMLE for $Y \Rightarrow$ targeted Q_T^*.
2. **Time** $T-1$ Treat Q_T^* as the outcome; run TMLE for $E[Q_T^* \mid \bar{A}_{T-1}, \bar{L}_{T-1}] \Rightarrow Q_{T-1}^*$.
3. **Iterate** Repeat until $t = 0$; each step uses its own clever covariate H_t built from g_t.
4. **Plug-in ATE** $\hat{\psi}_{\text{LTMLE}} = \frac{1}{n} \sum_i \left(Q_0^*(1, W_i) - Q_0^*(0, W_i) \right)$.

Iterative Double Robustness

Each backward TMLE step is doubly robust: consistent if *either* Q_t *or* g_t is correct.

Result: even if at *every* time point one model is wrong, the final LTMLE remains consistent as long as the *other* model is correct at that point.

Why LTMLE Is State-of-the-Art

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

- **Iterative double robustness** — two chances per time t.
- **Efficiency** — reaches the semiparametric lower bound if all models are right.
- **Machine-learning ready** — supports Super Learner + CV-TMLE at each step.
- **Substitution estimator** — predictions stay in natural bounds (0–1, etc.).

Summary of Longitudinal Methods

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Method	**Core Idea**	**Main Weakness**
Sequential G-Comp	Work backward (ICE) or forward simulation; plug-in predictions	All outcome models must be correct
Sequential IPTW	Weight by full treatment history; fit marginal structural model	All treatment models must be correct; extreme weights
LTMLE	Backward recursion with a TMLE update at each step	More computation

Bottom line

For time-varying confounding, **LTMLE** gives the most robust and efficient inference while letting you harness modern machine-learning tools.

From Theory to Policy: The FDA Real-World Evidence

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The demand for rigorous causal inference is no longer just academic. Regulatory bodies are actively grappling with how to use observational data.

The Rise of Real-World Evidence (RWE)

The U.S. FDA has a major initiative to use **Real-World Data** (from electronic health records, insurance claims, etc.) to support regulatory decision-making. This is impossible without credible causal inference.

TMLE's Role

TMLE and the Targeted Learning framework provide a pre-specified, transparent, and doubly-robust roadmap to analyze this messy data. Applications include:

- Post-market drug and device safety studies.
- Comparative effectiveness of different treatments.
- Augmenting RCTs with observational data to understand effects in a broader population.

Impact: Having a method like TMLE allows researchers to present findings to regulatory bodies with a higher degree of confidence, as the method is designed to minimize bias and be honest about its assumptions.

From Theory to Policy: The FDA Real-World Evidence

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The demand for rigorous causal inference is no longer just academic. Regulatory bodies are actively grappling with how to use observational data.

The Rise of Real-World Evidence (RWE)

The U.S. FDA has a major initiative to use **Real-World Data** (from electronic health records, insurance claims, etc.) to support regulatory decision-making. This is impossible without credible causal inference.

TMLE's Role

TMLE and the Targeted Learning framework provide a pre-specified, transparent, and doubly-robust roadmap to analyze this messy data. Applications include:

- Post-market drug and device safety studies.
- Comparative effectiveness of different treatments.
- Augmenting RCTs with observational data to understand effects in a broader population.

Impact: Having a method like TMLE allows researchers to present findings to regulatory bodies with a higher degree of confidence, as the method is designed to minimize bias and be honest about its assumptions.

From Theory to Policy: The FDA Real-World Evidence

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The demand for rigorous causal inference is no longer just academic. Regulatory bodies are actively grappling with how to use observational data.

The Rise of Real-World Evidence (RWE)

The U.S. FDA has a major initiative to use **Real-World Data** (from electronic health records, insurance claims, etc.) to support regulatory decision-making. This is impossible without credible causal inference.

TMLE's Role

TMLE and the Targeted Learning framework provide a pre-specified, transparent, and doubly-robust roadmap to analyze this messy data. Applications include:

- Post-market drug and device safety studies.
- Comparative effectiveness of different treatments.
- Augmenting RCTs with observational data to understand effects in a broader population.

Impact: Having a method like TMLE allows researchers to present findings to regulatory bodies with a higher degree of confidence, as the method is designed to minimize bias and be honest about its assumptions.

Recent Frontiers: High-Impact Applications

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Targeted Learning isn't just a theoretical exercise; it has been central to answering some of the most critical public health questions of our time.

COVID-19 Vaccine Efficacy

Problem: In the original COVID-19 vaccine trials, some participants "unblinded" themselves or didn't perfectly adhere to the protocol. How do you estimate the pure biological efficacy? **Solution:** Researchers from UC Berkeley, the FDA, and other institutions applied LTMLE to the trial data to adjust for these post-randomization behaviors, providing a more accurate estimate of the vaccine's true effect. (e.g., Luedtke et al., 2021)

Precision Medicine

Problem: Does a treatment work best for a specific subgroup of patients (e.g., based on their genetic markers)? **Solution:** TMLE can estimate treatment effects within data-adaptively defined subgroups, providing a path toward personalized medicine while maintaining valid statistical inference.

Recent Frontiers: High-Impact Applications

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Targeted Learning isn't just a theoretical exercise; it has been central to answering some of the most critical public health questions of our time.

COVID-19 Vaccine Efficacy

Problem: In the original COVID-19 vaccine trials, some participants "unblinded" themselves or didn't perfectly adhere to the protocol. How do you estimate the pure biological efficacy? **Solution:** Researchers from UC Berkeley, the FDA, and other institutions applied LTMLE to the trial data to adjust for these post-randomization behaviors, providing a more accurate estimate of the vaccine's true effect. (e.g., Luedtke et al., 2021)

Precision Medicine

Problem: Does a treatment work best for a specific subgroup of patients (e.g., based on their genetic markers)? **Solution:** TMLE can estimate treatment effects within data-adaptively defined subgroups, providing a path toward personalized medicine while maintaining valid statistical inference.

Final Takeaway: The Targeted Learning Mandate

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The core principle of Targeted Learning is a mandate for modern science:

"State the scientific question first, then use the best available data and the most flexible, robust statistical methods to provide an answer, while being transparent about all assumptions."

Your Role as a Public Health Leader

As the next generation of researchers and policymakers, embracing this framework is not just a statistical choice—it is a commitment to producing the most reliable and impactful evidence possible.

Final Takeaway: The Targeted Learning Mandate

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

The core principle of Targeted Learning is a mandate for modern science:

"State the scientific question first, then use the best available data and the most flexible, robust statistical methods to provide an answer, while being transparent about all assumptions."

Your Role as a Public Health Leader

As the next generation of researchers and policymakers, embracing this framework is not just a statistical choice—it is a commitment to producing the most reliable and impactful evidence possible.

Thank You

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Questions?

Contact:
David McCoy
david_mccoy@berkeley.edu
Please leave a review of the course

Start Here – Foundations & Tutorials

Core Papers

- Schuler MS, Rose S. (2017). Targeted maximum likelihood estimation for causal inference in observational studies. *Am. J. Epidemiol.* 185(1):65-73. DOI: 10.1093/aje/kww165
- Luque-Fernandez MA, Schomaker M, Rachet B, Schnitzer ME. (2018). TMLE for a binary treatment: a tutorial. *Stat. Med.* 37(16):2530-2546. DOI: 10.1002/sim.7628
- van der Laan MJ, Rubin D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* 2(1). DOI: 10.2202/1557-4679.1043

Books

- van der Laan MJ, Rose S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer. ISBN 978-1-4419-9781-4.
- van der Laan MJ, Rose S. (2018). *Targeted Learning in Data Science.* Springer. ISBN 978-3-319-65303-7.

Online Tutorials

- TLverse Learning Hub – free workshops, slides, and code: `https://tlverse.org`

Hands-On – Software & Implementation

R Packages

- Gruber S, van der Laan MJ. (2012). `tmle`: Targeted Maximum Likelihood Estimation. *J. Stat. Softw.* 51(13). DOI: 10.18637/jss.v051.i13
- Lendle SD et al. (2017). `ltmle`: TMLE for longitudinal data. *J. Stat. Softw.* 81(1). DOI: 10.18637/jss.v081.i01

tlverse Ecosystem

- `tmle3`, `sl3`, `origami` – modern modular TMLE in R.
  Docs: `https://tlverse.org`
- Interactive notebooks: `https://tlverse.org/tl-workshop`

Getting Started Quickly

- Vignettes in each package (run `browseVignettes("tmle3")` in R).
- Minimal Python: `zEpid` (see TLverse site for bridge tutorials).

Dig Deeper – Extensions & Applications

TMLE Tutorial

David McCoy PhD, MSc

The Causal Question

Traditional Approaches & Their Limits

TMLE: A Smarter, Targeted Approach

The Theory Behind the Targeting Step

Summary

TMLE for Longitudinal Data

Impact and Future Directions

Advanced Methods

- Petersen ML et al. (2014). TMLE for longitudinal marginal structural models. *J. Causal Inference* 2(2):147-185. DOI: 10.1515/jci-2013-0007
- van der Laan MJ, Luedtke AR. (2015). TMLE for optimal dynamic treatment rules. *J. Causal Inference* 3(1):61-95. DOI: 10.1515/jci-2013-0022
- Díaz I et al. (2023). Non-parametric causal effects via modified treatment policies. *JASA* 118(541):846-857. DOI: 10.1080/01621459.2021.1955691

Clinical Public-Health Applications

- Balzer LB et al. (2016). Adaptive pre-specification in randomized trials. *Stat. Med.* 35(25):4528-4545. DOI: 10.1002/sim.6991
- Benkeser D et al. (2021). Covariate adjustment in COVID-19 trials. *Biometrics* 77(4):1467-1481. DOI: 10.1111/biom.13377

Machine Learning Integration

- van der Laan MJ. (2017). Generally efficient TMLE via Highly Adaptive Lasso. *Int. J. Biostat.* 13(2). DOI: 10.1515/ijb-2015-0097

Systematic Review

- Smith MJ et al. (2023). Applications of TMLE in epidemiology: a systematic review. *Ann. Epidemiol.* 86:34-48. DOI: 10.1016/j.annepidem.2023.06.004

Journal of Causal Inference – Top TMLE Papers

- Baumann PFM, Schomaker M, Rossi E. (2021). Estimating the effect of central bank independence on inflation using longitudinal TMLE. *JCI* 9(1): 109-146.
  DOI: 10.1515/jci-2020-0016
- Petersen ML, Schwab J, Gruber S, et al. (2014). TMLE for dynamic and static longitudinal MSMs. *JCI* 2(2): 147-185.
  DOI: 10.1515/jci-2013-0007
- Luedtke AR, van der Laan MJ. (2015). Targeted learning of the mean outcome under an optimal dynamic treatment rule. *JCI* 3(1): 61-95.
  DOI: 10.1515/jci-2013-0022
- Athey S, Imbens GW, Wager S. (2019). A fundamental measure of treatment-effect heterogeneity (includes CV-TMLE estimator). *JCI* 7(1): 21-46.
  DOI: 10.1515/jci-2019-0003

Biometrics Biostatistics – Survival / Longitudinal TMLE

Key Survival Extensions

- Carone M, Saad N, Luedtke AR, van der Laan MJ. (2020). One-step TMLE for time-to-event outcomes. *Biometrics* 76(3): 889-899. DOI: 10.1111/biom.13172
- Rytgaard HCW, Eriksson F, van der Laan MJ. (2023). Time-specific intervention effects with continuous survival times via TMLE. *Biometrics* 79: 3038-3049. DOI: 10.1111/biom.13634

Highly Adaptive Lasso (HAL) Meets TMLE

- Díaz I, Luedtke AR, Carone M, van der Laan MJ. (2021). Targeted estimation for right-censored data using HAL. *Biostatistics* 22(4): 897-913. DOI: 10.1093/biostatistics/kxaa017

Clinical-Trial Applications Optimal Treatment Rules

Covariate-Adjusted Adaptive Trials

- Benkeser D, et al. (2021). Precision power in COVID-19 trials using TMLE covariate adjustment. *Biometrics* 77(4): 1467-1481. DOI: 10.1111/biom.13377
- Balzer LB, et al. (2016). Adaptive pre-specification in randomized trials. *Stat. Med.* 35(25): 4528-4545. DOI: 10.1002/sim.6991

Optimal Personalized Interventions

- Luedtke AR, van der Laan MJ. (2016). Optimum dynamic treatment rules via TMLE. (See JCI slide.)
- Laber EB, Porter KE, et al. (2018). Robust value-function estimation with TMLE in adaptive treatment-strategies research. *JRSS-B* 80(5): 1001-1020. DOI: 10.1111/rssb.12275